A new smoothed quasi maximum likelihood estimator for autoregressive process with LARCH errors.

Lionel Truquet

Abstract

We introduce a smoothed version of the quasi maximum likelihood estimator (QMLE) in order to fit heteroschedastic time series with possibly vanishing conditional variance. We apply this procedure to a finite-order autoregressive process with linear ARCH errors. We prove both the almost sure consiistency and the asymptotic normality of our estimator. This estimator is more robust that QMLE with the same type of assumptions. A numerical study confirms the qualities of our procedure.

Primary AMS keyword: 62M10 Inference for time series
Keywords: 62F12: Asymptotic properties of estimators. 60G10: Stationary processes

1 Introduction

In order to study the behaviour of financial time series such as asset returns or exchange rates, a considerable work has been done to study ARCH models introduced by Engle (1982) [7]. From empirical observations of time series, Black (1976) [2] called the leverage effect a tendency for the conditional variance to be negatively correlated with the past returns. Another property is a slow decay of the autocorrelations of the squares called long memory, see Doukhan *et al.* [5]. LARCH(∞) (Linear ARCH(∞)) models introduced in Giraitis *et al.* (2000) [12] and (2004) [13] take in account these two properties; they are defined from an equation:

$$X_t = \xi_t\Big(a_0 + \sum_{j\geq 1} a_j X_{t-j}\Big) \tag{1}$$

for an independent and identically distributed (i.i.d) sequence (ξ_t) with $\mathbb{E}\xi_0 = 0$ and $\mathbb{E}\xi_0^2 = 1$. Long memory properties of the model are addressed by Giraitis *et al.* (2000) [12] whereas the leverage property is studied in Giraitis *et al.* (2004) [13]. The model (1) specializes to the asymmetric ARCH model of Engle (1990) [8]. The conditional variance of models (1) writes as the square of a linear combination of the past values:

$$V_t = \Big(a_0 + \sum_{j\geq 1} a_j X_{t-j}\Big)^2$$

A short memory version of (1) is LARCH(p), here $a_j = 0$ for $j > p$ hence:

$$X_t = \xi_t \Big(a_0 + \sum_{j=1}^{p} a_j X_{t-j}\Big) \tag{2}$$

Then the model (2) is a special case of a more general model introduced in Sentena (1995) [18]; here the conditional variance writes as a quadratic form:

$$V_t = a_0 + \sum_{j=1}^{p} a_j X_{t-j} + \sum_{j=1}^{p}\sum_{k=1}^{p} b_{j,k} X_{t-j} X_{t-k} \tag{3}$$

If this quadratic form is nonnegative this is possible to exhibit assumptions ensuring $V_t \geq 0$; a solution (X_t) with this conditional variance writes $X_t = \xi_t V_t$ for iid inputs ξ_t. Giraitis *et al.* (2000) [12] prove a necessary and sufficient condition for the existence and the uniqueness of a square integrable and strictly stationary solution of the equation (1). Sufficient conditions for the existence of higher moments is also provided, moreover Giraitis *et al.* (2004) [13] explicit sufficient conditions for the leverage property.

The model (1) is generalized in Giraitis and Surgailis (2002) [14] to a bilinear model which exhibits long memory both in conditional mean and in conditional variance:

$$X_t = \alpha + \sum_{j\geq 1} \alpha_j X_{t-j} + \xi_t \Big(a_0 + \sum_{j\geq 1} a_j X_{t-j}\Big) \tag{4}$$

In the short memory case, Francq *et al.* (2008) [10] study existence and uniqueness of a strictly stationary solution of equation (2) (not necessarily square integrable).

A main statistical problem is to estimate the parameter $\theta = (a_0, \ldots, a_p)$ of the model (2). A classical estimation procedure is the Gaussian Quasi Maximum Likelihood Estimation (QMLE). Under conditions, the QMLE is shown to be consistent and asymptotically normal. But a crucial condition in its application is the existence of a real number $h > 0$ such that $V_0(\theta) \geq h$ a.s. For the model (2), the conditional variance V_0 is, in general, not bounded away from 0 and the quasi likelihood becomes numerically intractable. Because the QMLE cannot be used for the model (2), we propose a smoothed version of the QMLE which is more robust then the classical QMLE and applies with the same kind of assumptions. We apply this procedure to an AR process with LARCH errors.

The paper is organized as follows. Section 2 recalls the properties of the model (2). The next Section 3 introduces our model and mand motivates the introduction of our smoothed QMLE. Section 4 addresses its asymptotic propertiesfor our model. In Section 5, we discuss the behaviour of its asymptotic variance when the smoothing parameter tends to 0. Section 6 is dedicated to a numerical illustration. The proofs are postponed to the last section of the paper.

2 Some general results about LARCH models.

The first results about existence of LARCH models were given in the general case of equation (1). The condition $\sum_{j=1}^{\infty} a_j^2 < 1$ is necessary and sufficient for the existence of a square integrable and nonanticipative solution (see Theorem 2.1 in Giraitis *et al.* (2004) [13]). Those authors prove that the unique solution of equation (1) is defined from the Volterra expansion:

$$X_t = a_0 \xi_t \Big(1 + \sum_{k \geq 1} \sum_{j_1, \ldots, j_k \geq 1} a_{j_1} \cdots a_{j_k} \xi_{t-j_1} \cdots \xi_{t-(j_1+\cdots+j_k)} \Big) \tag{5}$$

Those authors also give a sufficient condition for the existence of the fourth moments in the general case of model (1):

$$\mu_4 \sum_{j=1}^{\infty} a_j^4 + 4 |\mu_3| \sum_{j=1}^{\infty} |a_j|^3 + 6 \sum_{j=1}^{\infty} a_j^2 < 1 \tag{6}$$

where $\mu_i = \mathbb{E}\xi_0^i$ for $1 \leq i \leq 4$ (here $\mu_2 = 1$). Mention that

$$\mu_4^{1/4} \sum_{j=1}^{\infty} |a_j| < 1 \tag{7}$$

ensures the existence of the fourth moment for the solution (5) (see Doukhan *et al.* (2006) [6]), hence the condition (6) is perhaps not sharp. Although condition (6) is less restrictive with respect to the decay of the sequence $(a_j)_{j \geq 1}$, condition (7) can obviously be better, *e.g.* if $a_j = 0$, $j \geq 2$.

From now on we fix an integer $p \geq 1$ and we only consider eqn. (2). Then, existence and uniqueness of a strictly stationary solution of (2) holds under the less restrictive condition $\sum_{j=1}^{p} a_j^2 < 1$, is pointed in Francq *et al.* (2008) [10]. Denote

$$A_t = \begin{pmatrix} a_{1:p-1}\xi_t & a_p \xi_t \\ I_{p-1} & 0_{p-1} \end{pmatrix}, \qquad \text{where} \qquad a_{1:p-1} = (a_1, \ldots, a_{p-1})$$

and I_k is the $k \times k$ identity matrix. If $p = 1$ then $A_t = a_1 \xi_t$. Let $A = (A_t)_t$ and $\gamma(A)$ the top-Lyapunov exponent of the sequence A:

$$\gamma(A) = \lim_{t \to \infty} \frac{1}{t} \log \|A_t \cdots A_1\|$$

Theorem 3.1 in Francq *et al.* (2008) [10] asserts that equation (2) admits a strictly stationary solution if and only if

$$\gamma(A) < 0. \tag{8}$$

Under this condition, the strictly stationary solution is unique, nonanticipative and ergodic.
If $p = 1$ condition (8) writes explicitly $|a_1| < e^{-\mathbb{E}\log|\xi_0|}$. As pointed in [10], if $\xi_0 \sim \mathcal{N}(0,1)$ this writes $|a_1| < 1.88736$. In comparison the condition for the existence of a second moment writes $|a_1| < 1$ and the one for the fourth moment $|a_1| < 0.7598\cdots$ (see section 3).

3 Model specification and smoothed QMLE

We consider for $p, q \in \mathbb{N}^*$ the model

$$\begin{aligned} Y_t &= b_{0,1}Y_{t-1} + \cdots + b_{0,q}Y_{t-q} + X_t, && (9)\\ X_t &= \xi_t \left(a_{0,0} + \sum_{j=1}^{p} a_{0,j}X_{t-j} \right), \qquad t \in \mathbb{Z} && (10) \end{aligned}$$

with ξ an i.i.d sequence such that $\mathbb{E}\xi_0 = 0$, $\mathbb{E}\xi_0^2 = 1$. By convention, $q = 0$ means that the process Y is a pure LARCH model given by (10); note that for $p = 0$ the model is an AR(q) process.
In the sequel, when we consider a solution of equation (9) or (10), it is always assumed that this solution is stationary, ergodic and non anticipative. We denote

$$\theta_0 = (b_{0,1}, \ldots, b_{0,q}, a_{0,0}, \ldots, a_{0,p}),$$

and for $\theta = (b_1, \ldots, b_q, a_0, \ldots, a_p) \in \mathbb{R}^{p+q+1}$ and $t \in \mathbb{Z}$:

$$\begin{aligned} m_t(\theta) &= \sum_{j=1}^{q} b_j Y_{t-j},\\ V_t(\theta) &= \sigma_t^2(\theta) = \left(a_0 + \sum_{j=1}^{p} a_j \left(Y_{t-j} - m_{t-j}(\theta)\right) \right)^2. \end{aligned}$$

Setting $\mathcal{F}_t = \sigma(Y_{t-1}, Y_{t-2}, \ldots)$ for $t \in \mathbb{Z}$, we have:

$$m_t(\theta_0) = \mathbb{E}\left(Y_t / \mathcal{F}_{t-1}\right), \qquad V_t(\theta_0) = \mathrm{Var}\,\left(Y_t / \mathcal{F}_{t-1}\right).$$

By stationarity, we can always suppose that the data $Y_n, Y_{n-1}, \ldots, Y_{-(p+q)+1}$ are available. Usually, the QMLE is defined by:

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} Q_n(\theta),$$

$$Q_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \frac{(Y_t - m_t(\theta))^2}{V_t(\theta)} + \ln V_t(\theta). \tag{11}$$

Although we will not prove any result about the consistency or the inconsistency of the QMLE for the model (9), it seems very difficult to compute this estimator because of the intractable form of the function $\theta \mapsto Q_n(\theta)$ (see figure 1 for which $q = 0$, $p = 1$, $a_0 = 1$; the data are generated with $a_1 = 0.5$ and $\xi_0 \sim \mathcal{N}(0,1)$). The roughness of the function $\theta \mapsto Q_n(\theta)$ is due to the small values of the function $\theta \mapsto V_t(\theta)$; this gives infinite values for Q_n.

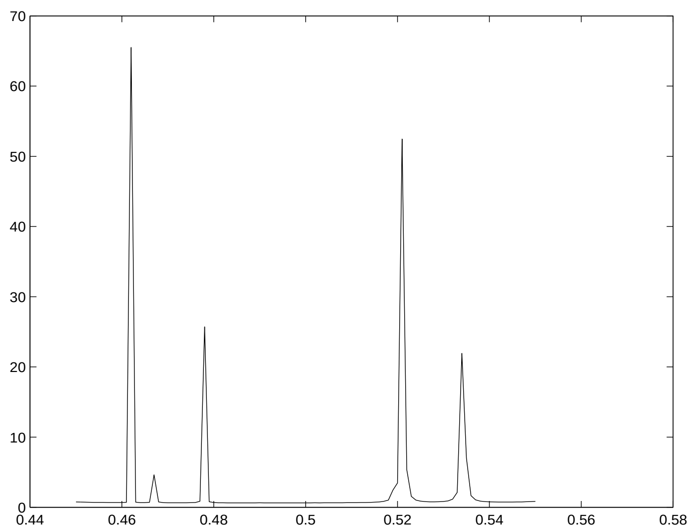

Figure 1: $a_1 \mapsto Q_{500}(\theta)$.

In some cases, even the conditional variance $V_0(\theta_0)$ is not bounded away from zero as shows the following Lemma:

Lemma 1 *Suppose that for the model (10) the input ξ_0 admits a density with support $\mathbb{R}$ and $a_{0,0} \neq 0$. If $j_0 = \min\{j / a_{0,j} \neq 0\}$ exists, then the conditional variance $V_0(\theta_0)$ may be unbounded away from zero.*

One also can note that even if the conditonal variance is not bounded away from zero, it does not vanish if ξ_0 is atomless:

Lemma 2 *Suppose that for model (10), the law of ξ_0 is atomless, then $\mathbb{P}(\sigma_0(\theta_0) = 0) = 0$.*

We define here a new contrast working also will small values of the conditional variance. Suppose just for a moment (even this is not true) that (Y_t) is a ARCH process with a conditional variance bounded away from zero; if

$\xi_0 \sim \mathcal{N}(0,1)$ then $-1/2(Q_n + \ln 2\pi)$ is well defined and is the exact conditional log-likelihood of the random vector $(Y_1, \ldots, Y_n)$. For nonGaussian inputs, ξ_0, this is not true anymore but the function

$$Q : \theta \mapsto \mathbb{E}\left(\frac{(Y_0 - m_0(\theta))^2}{V_0(\theta_0)} + \ln(V_0(\theta)) \right) \tag{12}$$

is still a good contrast since θ_0 is the unique minimum of Q, provided the following identification condition holds:

$$(m_0(\theta), V_0(\theta)) = (m_0(\theta_0), V_0(\theta_0)) \Rightarrow \theta = \theta_0.$$

In fact, Q can be used as a contrast for the estimation of a parameter of the mean and/or the variance of a conditional law W_0/U_0, for some stationary ergodic process $\{(U_t, W_t)/t \in \mathbb{Z}\}$; in our case $U_t = \big(Y_{t-1}, \ldots, Y_{t-(p+q)}\big)$ and $W_t = Y_t$.

Set now $U_t = \big(Y_{t-1}, \ldots, Y_{t-(p+q)}\big)$ and $W_t = Y_t + \eta_t$, for $t \in \mathbb{Z}$, where $(\eta_t)_t$ denotes an i.i.d sequence, independent of the process (Y_t), with $\mathbb{E}\eta_0 = 0$ and $\mathrm{Var}\,(\eta_0) = h$ for some real number $h > 0$. Then we have

$$\mathbb{E}\,(W_t/U_t) = m_t(\theta_0), \qquad \mathrm{Var}\,(W_t/U_t) = V_t(\theta_0) + h,$$

and the contrast Q becomes

$$Q_h(\theta) = \mathbb{E}\left(\frac{(W_0 - m_0(\theta))^2}{V_0(\theta) + h} + \ln(V_0(\theta) + h) \right).$$

We obtain from independence,

$$Q_h(\theta) = \mathbb{E}\left(\frac{(Y_0 - m_0(\theta))^2 + h}{V_0(\theta) + h} + \ln(V_0(\theta) + h) \right).$$

The number $h > 0$ avoids the problem of small possible values for the variance in (12): it will be called the *smoothing parameter*. If the data $Y_n, Y_{n-1}, \ldots, Y_{-(p+q)+1}$ are available, we define the following estimator:

$$\begin{aligned} \hat{\theta}_{n,h} &= \arg\min_{\theta \in \Theta} Q_{n,h}(\theta), && (13) \\ Q_{n,h}(\theta) &= \frac{1}{n} \sum_{t=1}^{n} q_{t,h}(\theta), && (14) \\ q_{t,h}(\theta) &= \frac{(Y_t - m_t(\theta))^2 + h}{V_t(\theta) + h} + \ln\left(V_t(\theta) + h\right). && (15) \end{aligned}$$

Observe that $\hat{\theta}_{n,0}$ is the classical QMLE. For $h > 0$ and $n \in \mathbb{N}^*$, $Q_{n,h}$ has a more tractable expression that $Q_{n,0}$. The asymptotic properties of the estimator $\hat{\theta}_{n,h}$, called *smoothed QMLE*, will be derived below.

4 Asymptotics of smoothed QMLE for AR-LARCH models

QMLE is very popular for conditionally heteroscedastic time series. Its asymptotic properties were first established by Weiss (1986) [20] for ARCH models. General results for the consistency of this method are proved in Jeantheau (1998) [11]. Both its consistency and its asymptotic normality are precised by Mikosch and Straumann (2006) [17] who set a nice theoretical framework for the univariate case. For multivariate time series we defer the reader to Bardet and Wintenberger (2007) [1]. For GARCH models, mention among others the works of Lee and Hansen (1994) [15], Lumsdaine (1996) [16], Berkes, Horváth and Kokoszka (2003) [3] and Francq and Zakoïan (2004) [9]. As we will see, asymptotics properties of the smoothed QMLE can be obtained using the same arguments as for the classical QMLE.
Let us introduce some assumptions:

(A1): $\gamma(A(\theta_0)) < 0$.

(A2): The roots of the polynomial P defined by $P(z) = 1 - \sum_{j=1}^{q} b_{0,j} z^j$ are outside the unit disk.

(A3): $\theta_0 \in \Theta$, a compact set such as for all $\theta \in \Theta$, the first component a_0 of θ is strictly positive.

(A4): The support of the law of ξ_t admits more than 2 points.

(A5): θ_0 belongs to the interior Θ° of Θ.

(A6): $\mathbb{E}X_0^4 < \infty$.

The top-Lyapounov exponent $\gamma(A(\theta_0))$ is defined for the LARCH part only, as in (8). Assumptions **(A1)** and **(A2)** ensure existence and uniqueness of the AR-LARCH process (9). The two following results are devoted respectively to *a.s.* consistency and to the central limit behaviour of the smoothed QMLE.

Theorem 1 *Under assumptions* $(\mathbf{A1}) - (\mathbf{A4})$ *the smoothed QMLE is consistent for each value of* $h > 0$*:*

$$\lim_{n\to\infty} \hat{\theta}_{n,h} = \theta_0, \qquad a.s.$$

Theorem 2 *If* **(A1)**-**(A6)** *hold true, the smoothed QMLE is asymptotically normal for each value of* $h > 0$*:*

$$\sqrt{n}\left(\hat{\theta}_{n,h} - \theta_0\right) \xrightarrow{\mathcal{D}}_{n\to\infty} \mathcal{N}\left(0, N_h^{-1} M_h N_h^{-1}\right).$$

where

$$
\begin{aligned}
N_h &= N_h^{(1)} + N_h^{(2)}, \quad M_h = M_h^{(1)} + M_h^{(2)} + M_h^{(3)}, \\
N_h^{(1)} &= 2\mathbb{E}\left(\frac{\nabla m_0(\theta_0)\nabla m_0(\theta_0)'}{V_0(\theta_0)+h}\right), \quad N_h^{(2)} = \mathbb{E}\left(\frac{\nabla V_0(\theta_0)\nabla V_0(\theta_0)'}{(V_0(\theta_0)+h)^2}\right), \\
M_h^{(1)} &= 4\mathbb{E}\left(\frac{V_0(\theta_0)\nabla m_0(\theta_0)\nabla m_0(\theta_0)'}{(V_0(\theta_0)+h)^2}\right), \\
M_h^{(2)} &= (\mu_4-1)\mathbb{E}\left(\frac{V_0(\theta_0)^2\nabla V_0(\theta_0)\nabla V_0(\theta_0)'}{(V_0(\theta_0)+h)^4}\right), \\
M_h^{(3)} &= 2\mu_3\mathbb{E}\left(\frac{V_0(\theta_0)\sigma_0(\theta_0)}{(V_0(\theta_0)+h)^3}\left(\nabla m_0(\theta_0)\nabla V_0(\theta_0)' + \nabla V_0(\theta_0)\nabla m_0(\theta_0)'\right)\right).
\end{aligned}
$$

Remark. If $q=0$ then Y is a pure LARCH model (10) and we obtain the consistency and the asymptotic normality of the smoothed QMLE as above. Its asymptotic variance writes as $\big(N_h^{(2)}\big)^{-1} M_h^{(2)} \big(N_h^{(2)}\big)^{-1}$.

5 Choice of the smoothing parameter h

We aim here at precising the asymptotic variance of the smoothed QMLE. Although we have proved the asymptotic properties of the smoothed QMLE only for the AR-LARCH model (9) the following study applies to more general heteroschedastic time series

$$X_t = m_\theta(X_{t-1},\dots,X_{t-q}) + \sigma_\theta(X_{t-1},\dots,X_{t-(p+q)})\xi_t$$

however, for such models, as in Bardet and Wintenberger (2008), the problem would be to check identifiability conditions. We denote by $\|\cdot\|$ the Euclidean norm for a vector or a matrix. For simplicity we write m (resp. V) instead of $m_0(\theta_0)$ (resp. $V_0(\theta_0)$) and $\nabla m, \nabla V$ for the gradient vectors. Using the notations in Theorem 2 we denote $v_h = N_h^{-1} M_h N_h^{-1}$ the asymptotic variance of the smoothed QMLE (see Theorem 2).
Unexpected results appear by plotting the asymptotic variance of the smoothed QMLE for small values of h. Suppose that we want to estimate the parameter a of the model: $X_t = \xi_t\,(1 + aX_{t-1})$ where $\xi_0 \sim \mathcal{N}(0,1)$. Then the asymptotic variance of the smoothed QMLE denoted by $v_h(a)$ seems to verify $\lim_{h\to 0} v_h(a) = 0$ for a large subset of parameters (see figure 2).

To study the behaviour of the asymptotic variance we set $A \preceq B$, the relation of order between symmetric positive definite matrices such that $x'Ax \le x'Bx$ if for each $x \in \mathbb{R}^d$, here A and $B \in \mathcal{M}_d(\mathbb{R})$. We will use the notation $A \prec B$ if $x'Ax < x'Bx$ for all $x \in \mathbb{R}^d \setminus \{0\}$.
In the following Lemma we discuss the qualitative behaviour of $h \mapsto v_h$. Even if we were not able to check monotonicity of this function (with the order $\preceq$), we shall precise the behaviour of the asymptotic variance at the

origin: here $v = \lim_{h\to 0^+} v_h = \inf_{h>0} v_h$ is either degenerated or has the same form that the asymptotic variance of the classical QMLE.

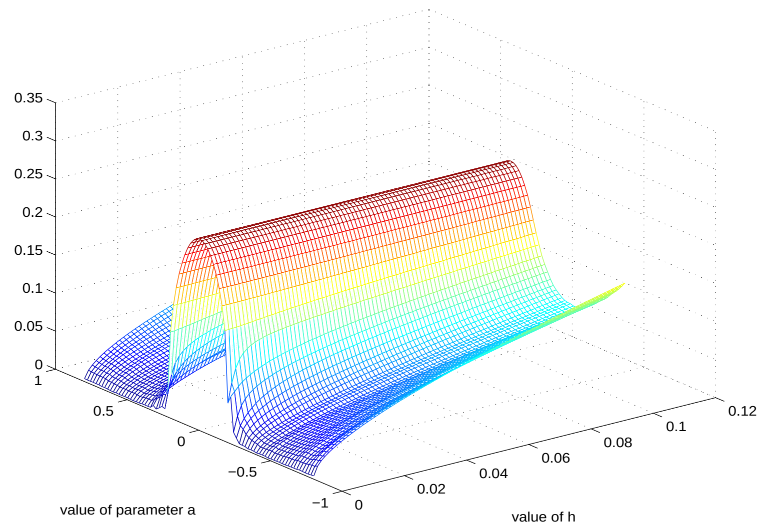

Figure 2: $(h, a) \mapsto v_h(a)$.

The behaviour of the asymptotic variance near $h = 0$ is related to the condition:

$$\mathcal{C}: \quad \mathbb{E}\left(\frac{\|\nabla m\|^2}{V}\mathbf{1}_{V\neq 0}\right) + \mathbb{E}\left(\frac{\|\nabla V\|^2}{V^2}\mathbf{1}_{V\neq 0}\right) < \infty.$$

Of course if $q = 0$ then the condition $\mathcal{C}$ reduces to

$$\mathbb{E}\left(\frac{\|\nabla V\|^2}{V^2}\mathbf{1}_{V\neq 0}\right) < \infty.$$

One can remark that when $\xi_0 \sim \mathcal{N}(0,1)$, by lemma 2 we have $V \neq 0$ a.s and condition $\mathcal{C}$ ensures the existence of the conditional Fisher information.

Lemma 3 *Let the assumptions of Theorem 2 hold. We suppose that either $q = 0$, or $q \neq 0$ but $(\mu_3, \mu_4) = (0, 3)$ (i.e ξ_0 has the same four first moments that a standard Gaussian). Then:*

1. $v = \lim_{h\to 0^+} v_h$ *exists, and* $v \preceq v_h$, *for* $h > 0$.
2. v *is non degenerated if and only if condition* $\mathcal{C}$ *holds. In this case,* $v = (\mu_4 - 1)N^{-1}$ *where:*

$$N = 2\mathbb{E}\left(\frac{\nabla m \nabla m'}{V}\mathbf{1}_{V\neq 0}\right) + \mathbb{E}\left(\frac{\nabla V \nabla V'}{V^2}\mathbf{1}_{V\neq 0}\right)$$

Remarks. **a.** Condition $\mathcal{C}$ holds if there exists $m > 0$ with:

$$V_0(\theta_0) \geq m > 0 \qquad a.s. \tag{16}$$

This is the case for example if ξ_0 has a uniform distribution on $[-\sqrt{3}, \sqrt{3}]$ and $a = \sum_{j=1}^{p} |a_{0,j}| < \frac{1}{2\sqrt{3}}$. Indeed we have $V_0(\theta_0) = a_{0,0}^2 \left(1 + \widetilde{\sigma}_0\right)^2$ where

$$\widetilde{\sigma}_0(\theta_0) = \sum_{k\geq 1} \sum_{j_1,\ldots,j_k \in \{1,\ldots,p\}} a_{0,j_1} \cdots a_{0,j_k} \xi_{-j_1} \cdots \xi_{-(j_1+\ldots+j_k)}.$$

Note that $|\widetilde{\sigma}_0(\theta_0)| \leq \frac{a\sqrt{3}}{1-a\sqrt{3}} < 1$ and $m = a_{0,0}^2 \left(1 - \frac{a\sqrt{3}}{1-a\sqrt{3}}\right)^2$ is a convenient value for (16) to hold. Condition (16) is a classical assumption to get the asymptotic properties of the classical QMLE, but for model (9), this kind of restriction seems unrealistic.

b. From point 2 in Lemma 3, if the condition $\mathcal{C}$ does not hold, then no asymptotically efficient estimator with $\sqrt{n}$-rate can be exhibited. This is the case if:

$$\mathbb{E} V_0(\theta_0)^{-1} \mathbf{1}_{V_0(\theta_0)\neq 0} = \infty. \tag{17}$$

Condition (17) is related to the behaviour of $V_0(\theta_0)$ around 0. The following (artificial) example shows that this condition may hold.
Suppose that $b_{0,j} = 0$, $j = 1, \ldots, q$ and $p = 1$, $a_{0,0} = 1$, $a_{0,1} = 0.5$, $\mathbb{P}(\xi_0 = 1) = \mathbb{P}(\xi_0 = -1) = \alpha$ and $\mathbb{P}(\xi_0 = 0) = 1 - 2\alpha$ for $\alpha \in [1/4, 1/2)$.
Then $\mathbb{E}\xi_0 = 0$ and from (7), $\mathbb{E}X_0^4 < \infty$ and we may assume $\mathbb{E}\xi_0^2 = 1$, for this we write $X_t = \xi_t / \sqrt{\mathbb{E}\xi_0^2} \left(\sqrt{\mathbb{E}\xi_0^2} + 0.5\sqrt{\mathbb{E}(\xi_0^2)}\, X_{t-1}\right)$. From (5) the chaotic expansion of the solution writes:

$$X_t = \xi_t + \sum_{j\geq 1} 2^{-j} \xi_t \cdots \xi_{t-j}$$

Let $n \in \mathbb{N}^*$ and suppose that $\xi_t = -1, \xi_{t-1} = \cdots = \xi_{t-n} = 1$ and $\xi_{t-(n+1)} = 0$, then:

$$X_t = -2(1 - 2^{-(n+1)})$$

and thus $V_{t+1}(\theta_0) = 2^{-(2n+2)}$. We now deduce:

$$\begin{aligned}
\mathbb{E}\frac{\mathbf{1}_{V_{t+1}(\theta_0)\neq 0}}{V_{t+1}(\theta_0)} &\geq \sum_{n\geq 1} 2^{2n+2} \mathbb{P}\left(V_{t+1}(\theta_0) = 2^{-(2n+2)}\right) \\
&\geq \sum_{n\geq 1} 2^{2n+2} \mathbb{P}\left(\xi_t = -1, \xi_{t-1} = \cdots = \xi_{t-n} = 1, \xi_{t-(n+1)} = 0\right) \\
&= (1 - 2\alpha) \sum_{n\geq 1} \alpha^{n+1} 2^{2n+2} \\
&= \infty
\end{aligned}$$

This example shows that the condition (17) may happen to hold. Now figure 2 seems to prove that the model LARCH(1) also exhibits this condition for $\xi_0 \sim \mathcal{N}(0,1)$ but no formal proof is given here.

c. It is clear that both the QMLE and the smoothed QMLE apply for classical ARCH models because their conditional variance is bounded away from zero. Then under the assumptions of lemma 3 the point 2. of the Lemma shows that the QMLE is more efficient than the smoothed QMLE.

6 Numerical illustration

We illustrate the behaviour of the smoothed QMLE with an example. Our goal here is to see if $h \to 0$ gives best estimators as suggested by the Lemma 3. We set $p = q = 1$ and we consider Gaussian errors. We recall that asymptotic normality of the smoothed QMLE requires $\mathbb{E}Y_0^4 < \infty$ (Theorem 2); moreover $\mathbb{E}X_0^4 < \infty \Rightarrow \mathbb{E}Y_0^4 < \infty$. The following Lemma gives a necessary and sufficient condition for the existence of the fourth moment of the solution of

$$X_t = \xi_t \left(a_{0,0} + a_{0,1}X_{t-1}\right) \tag{18}$$

if $\mathbb{E}\xi_0^3 = 0$.

Lemma 4 *Suppose that $\mathbb{E}\xi_0^3 = 0$ then there exists a stationary solution of equation (18) with $\mathbb{E}X_0^4 < \infty$ if and only if $a_1^4\mathbb{E}\xi_0^4 < 1$. In this case this solution is the unique stationary solution of equation (18).*

Remarks.

- If $\xi_0 \sim \mathcal{N}(0,1)$ is a standard normal random variable the condition $a_{0,1}^4\mathbb{E}\xi_0^4 < 1$ writes $|a_{0,1}| < 3^{-1/4} \approx 0.7598\cdots$.
- If ξ_0 follows the uniform distribution law on the interval $[-\sqrt{3},\sqrt{3}]$ this condition writes $|a_{0,1}| < (5/9)^{1/4} \approx 0.8633\cdots$. Thus if $|a_{0,1}| > 1/\sqrt{3} \approx 0.5774\cdots$, the process X is not bounded.

For the simulation study we have computed 500 smoothed QML estimators for sample sizes $n = 100$ and $n = 1000$ and for the smoothing parameters $h = 0.5, 0.1$ and 0.001. The value of the parameter is $\theta_0 = (0.5, 1.6, -0.7)$. An expected problem is the irregularity of the function $Q_{n,h}$ when h is small. This holds even for very large values of n. As an example we plot $a_1 \mapsto Q_{n,h}(a_1)$ in figure 3 for the model:

$$X_t = \xi_t \left(1 + 0.5X_{t-1}\right), \qquad \xi_0 \sim \mathcal{N}(0,1).$$

Then, to avoid optimization problems, we first compute the estimators for $h = 0.5$; after this, using those values to initialize the procedure, we start with an optimization procedure for $h = 0.1, 0.001$. We see from figure 5 that the mean square errors decrease as soon as h decreases. However if h is small, fitting to a Gaussian distribution is not very good for $n = 100$ (figure 4 and figure 6) but a large sample size $n = 1000$ corrects this problem. Hence the choice of the value of $h = h_n$ (depending on the sample size n) seems crucial. This problem is beyond the scope of this paper because we did not exhibit a balance of terms explaining this phenomenon as this is usual *e.g.* for non-parametric estimation.

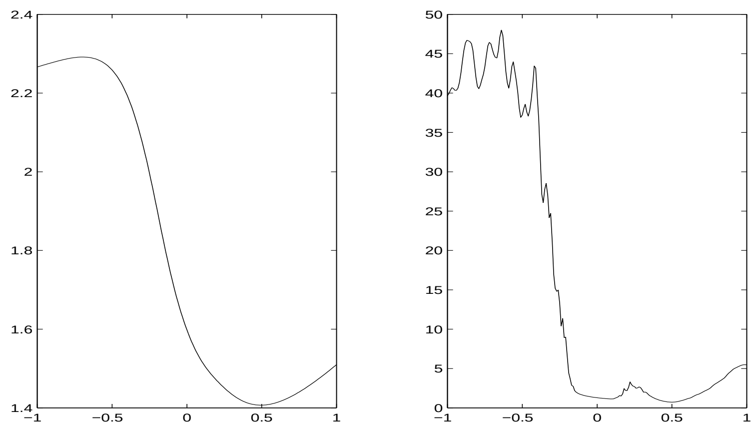

Figure 3: $Q_{n,h}$, $n = 20000$, $h = 0.5$ or $h = 0.001$.

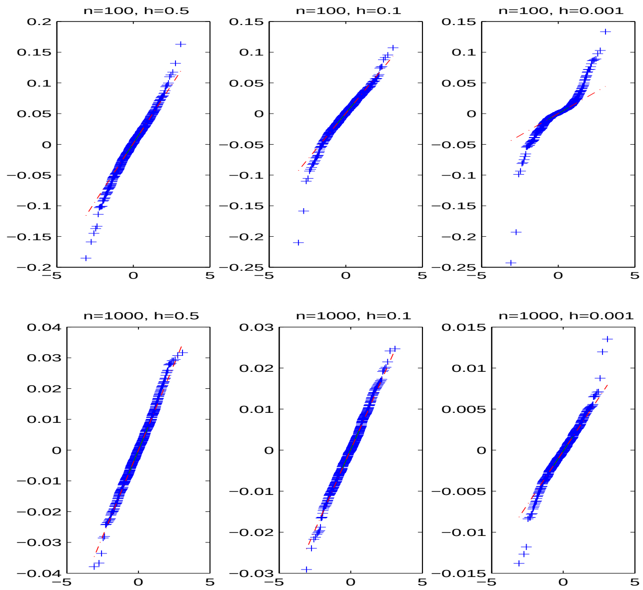

Figure 4: Normal Q-Q plots for the errors $\hat{b}_1 - b_{0,1}$.

Estimators	Sample size	S. QML		
		$h = 0.5$	0.1	0.001
$\hat{b}_1$	$n = 100$	1.9	1.1	0.9
$\hat{a}_0$	$n = 100$	16.3	14.7	13.9
$\hat{a}_1$	$n = 100$	8	5.8	5.1
$\hat{b}_1$	$n = 1000$	0.1	0.1	0
$\hat{a}_0$	$n = 1000$	1.7	1.5	1.4
$\hat{a}_1$	$n = 1000$	0.6	0.4	0.3

Figure 5: Mean square errors for the three estimators ($\times 10^{-3}$).

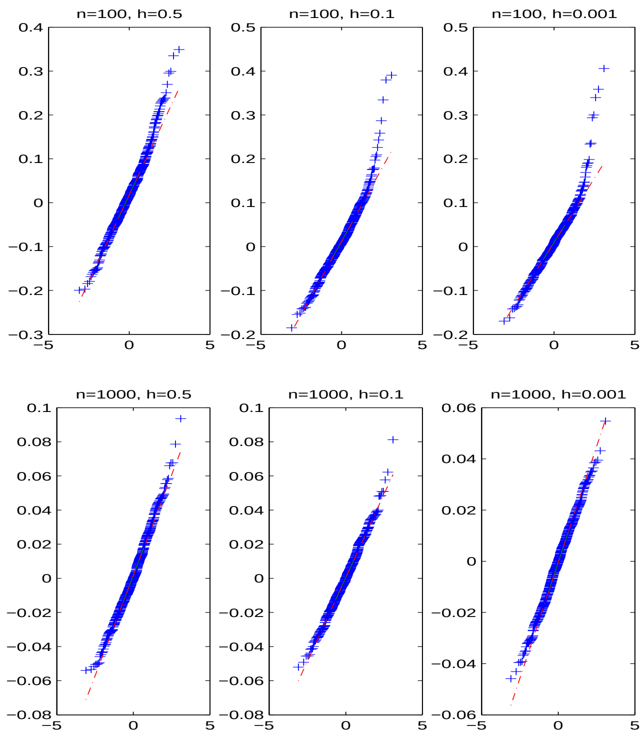

Figure 6: Normal Q-Q plots for the errors $\hat{a}_1 - a_{0,1}$.

7 Proofs

7.1 Proof of Lemma 1

Let $A = \{(z_1, \ldots, z_p) \in \mathbb{R}^p / a_{0,0} + \sum_{j=1}^{p} a_{0,j} z_j \neq 0\}$ and for $t \in \mathbb{Z}$, $Z_t = (X_{t-1}, \ldots, X_{t-p})$. Set $\sum_{j=j_0+1}^{p} = 0$ if $j_0 = p$. Then since $a_{0,0} \neq 0$, $X_t \equiv 0$ is not a solution of the equation (10) thus $\mathbb{P}(Z_t \in A) > 0$. Moreover for $z \in A$ the support of the conditional law $\mathcal{L}(X_{-j_0}/Z_{-j_0} = z)$ is the whole set $\mathbb{R}$ which is in contradiction with the existence of a number $m > 0$ such that

$$V_0(\theta_0) = \left(a_{0,0} + a_{j_0} X_{-j_0} + \sum_{j=j_0+1}^{p} a_j X_{-j}\right)^2 \geq m, \quad \text{a.s.} \quad \square$$

7.2 Proof of Lemma 2

For simplicity, we denote σ_t instead of $\sigma_t(\theta_0)$. The result is obvious if $a_{0,j} = 0$, $j = 1, \ldots, p$ since in this case $\sigma_0 = a_{0,0} \neq 0$.
Now let $j_0 \geq 1$ be the first index such that $a_{0,j_0} \neq 0$. Let $\alpha = \mathbb{P}(\sigma_0 = 0)$.

We prove by induction that:

$$\forall n \in \mathbb{N}, \quad \mathbb{P}\left(A_n\right) = \alpha \tag{19}$$

where for $n \in \mathbb{N}$, $A_n = \bigcap_{l=0}^{n} \{\sigma_{-lj_0} = 0\}$. This will conclude the proof. Indeed setting $n \to \infty$, we deduce that $\mathbb{P}\left(\bigcap_{l=0}^{\infty} \{\sigma_{-lj_0} = 0\}\right) = \alpha$ and from the ergodicity of the process $(\sigma_{-lj_0})_{l \in \mathbb{N}}$ we derive that $\alpha \in \{0, 1\}$. However $\alpha = 1$ implies by definition $\sigma_0 = X_0 = 0$ *a.s* which is impossible if $a_{0,0} \neq 0$: hence $\alpha = 0$.

We now prove (19). The definition of α implies the result for $n = 0$. Suppose that $\mathbb{P}(A_n) = \alpha$ and let us prove that $\mathbb{P}(A_{n+1}) = \alpha$. Then it is enough to prove $\mathbb{P}\left(A_n \cap \{\sigma_{-(n+1)j_0} \neq 0\}\right) = 0$ or:

$$\mathbb{P}\left(\sigma_{-nj_0} = 0, \sigma_{-(n+1)j_0} \neq 0\right) = \mathbb{P}\left(\sigma_0 = 0, \sigma_{-j_0} \neq 0\right) = 0. \tag{20}$$

Now $\sigma_0 = 0 \Leftrightarrow X_{-j_0} = \xi_{-j_0}\sigma_{-j_0} = M$, with $M = -\left(\sum_{j=j_0+1}^{p} a_{0,j} X_{-j}\right)/a_{0,j_0}$ (by convention $\sum_{j=p+1}^{p} = 0$). If μ is the law of the vector (σ_{-j_0}, M), we get using independence:

$$\mathbb{P}\left(\sigma_0 = 0, \sigma_{-j_0} \neq 0\right) = \int \mathbb{P}(a\xi_{-j_0} = b, a \neq 0)\mu(da, db).$$

Since $\mathbb{P}(\xi_0 = x) = 0$, $\forall x \in \mathbb{R}$, we have etablished (20), and (19) follows by induction. $\square$

7.3 Proof of Theorem 1

We first prove the following Lemma, useful to show that the parameter θ_0 is identifiable in the model (9).

Here Y is a model satisfying (9) and note that $\mathcal{F}_t = \sigma\left(X_{t-j}/j \in \mathbb{N}\right) = \sigma\left(Y_{t-j}/j \in \mathbb{N}\right)$ for $t \in \mathbb{Z}$.

Lemma 5 *We suppose that* **(A3)** *holds. Let U_1 and U_2 be two random variables measurable w.r.t $\mathcal{F}_{-1}$ and $(\alpha_j)_{0 \le j \le p}$ and $(\beta_j)_{0 \le j \le p}$ be real numbers such that $\beta_0 \neq 0$. Then*

1. $(X_0 - U_1) \times U_2 = 0$ *a.s* $\Rightarrow U_2 \times \sigma_0(\theta_0) = 0$ *a.s and* $U_1 \times U_2 = 0$ *a.s.*
2. $\mathbb{P}\left((X_0 - U_1) \times (X_0 - U_2) = 0\right) < 1$.
3. $\mathbb{P}\left(\sigma_0(\theta_0)\left(\beta + \sum_{j=1}^{p} \beta_j X_{-j}\right) = 0\right) < 1$.
4. $\left(\alpha_0 + \sum_{j=1}^{p} \alpha_j X_{-j}\right) \times \left(\beta_0 + \sum_{j=1}^{p} \beta_j X_{-j}\right) = 0$, *a.s implies* $\alpha_j = 0$, *for all* $j = 0, \ldots, p$.

Proof.

1. Here $(X_0 - U_1)\, U_2 = 0 \ a.s \Rightarrow U_2\sigma_0(\theta_0)\xi_0 = U_1U_2$. Since ξ_0 is not a constant and it is independent of $(U_2\sigma_0(\theta_0), U_1U_2)$ we have $U_2\sigma_0(\theta_0) = 0 \ a.s.$, thus obviously $U_1U_2 = 0 \ a.s.$

2. If $(X_0 - U_1)\,(X_0 - U_2) = 0 \ a.s$ then

$$\sigma_0^2(\theta_0)\xi_0^2 + \sigma_0(\theta_0)\,(-U_1 - U_2)\,\xi_0 = -U_1U_2 \quad a.s$$

   Since $a_{0,0} \neq 0$, we have $X \neq 0$ a.s. Then there exists a realization of $X_{-1}, X_{-2}, \ldots$ such that $\sigma_0(\theta_0) \neq 0$. For such a realization the support of the conditional law $\xi_0/X_{-1}, X_{-2}, \ldots$ (by independence this is also the law of ξ_0) contains only two values. This yields a contradiction with **(A3)** and the result follows.

3. We suppose

$$\sigma_0(\theta_0)\Big(\beta_0 + \sum_{j=1}^{p} \beta_j X_{-j}\Big) = 0, \quad a.s \tag{21}$$

   We set $\beta_j = 0$ for $j \geq p+1$. Suppose that $l = \inf\{i \geq 1/a_{0,i} \neq 0\}$ exists. We will show by induction that for all $i \in \mathbb{N}$:

$$\sigma_{-il}(\theta_0)\Big(\beta_0 + \sum_{j \geq il+1} \beta_j X_{-j}\Big) = 0, \qquad a.s.$$

   From (21), the result holds for $i = 0$. Suppose that for $i \in \mathbb{N}$:

$$\sigma_{-il}(\theta_0)\Big(\beta_0 + \sum_{j \geq il+1} \beta_j X_{-j}\Big) = 0 \qquad a.s.$$

   Then successive applications of point 1) lead to

$$\sigma_{-il}(\theta_0)\Big(\beta_0 + \sum_{j \geq (i+1)l} \beta_j X_{-j}\Big) = 0, \quad a.s.$$

   Now as $\sigma_{-il}(\theta_0) = a_{0,0} + \sum_{j=l}^{p} a_{0,j}X_{-il-j}$ and $a_{0,l} \neq 0$ we deduce from point 2) that $\beta_{(i+1)l} = 0$. Moreover a new application of point 1) leads to $\sigma_{-(i+1)l}(\theta_0)\left(\beta + \sum_{j \geq (i+1)l+1} \beta_j X_{-j}\right) = 0 \ a.s.$ Hence the result follows by induction.
   Now if i is large enough:

$$\sigma_{-il}(\theta_0)\beta_0 = 0 \quad a.s.$$

   Since $\beta_0 \neq 0$ this implies $\sigma_{-il}(\theta_0) = 0$ a.s. We have obtained a contradiction since X cannot equals 0 when $a_{0,0} \neq 0$.
   If now l does not exist, we have $X = a_{0,0}\xi$ and equation (21) becomes $a_{0,0}\left(\beta_0 + \sum_{j \geq 1} \beta_j a_{0,0}\xi_{-j}\right) = 0$ *(a.s)*. Taking expectations this equality leads to $\beta_0 = 0$ and we thus exhibit a contradiction.
   We have shown that relation (21) is not possible and the result follows.

4. Setting $\alpha_j = \beta_j = 0$ if $j \geq p+1$ we suppose $\left(\alpha_0 + \sum_{j\geq 1} \alpha_j X_{-j}\right) \times \left(\beta_0 + \sum_{j\geq 1} \beta_j X_{-j}\right) = 0$ *(a.s.)* An application of point 2) implies $\alpha_1 \times \beta_1 = 0$. Moreover an application of point 1) gives
$$\left(\alpha_0 + \sum_{j\geq 2} \alpha_j X_{-j}\right) \times \left(\beta_0 + \sum_{j\geq 2} \beta_j X_{-j}\right) = 0$$
Then by an induction argument, it is obvious that for $i \geq 1$, we will obtain $\alpha_i \beta_i = 0$ and
$$\left(\alpha_0 + \sum_{j\geq i} \alpha_j X_{-j}\right) \times \left(\beta_0 + \sum_{j\geq i} \beta_j X_{-j}\right) = 0 \quad a.s. \tag{22}$$
With $i \to \infty$ we thus derive $\alpha_0\beta_0 = 0$, hence $\alpha_0 = 0$. Suppose that there exists some $i \in \mathbb{N}^*$ such that $\alpha_i \neq 0$. Then $\beta_i = 0$ and applying point 1) to equality (22) we get
$$\sigma_{-i}(\theta_0)\left(\beta_0 + \sum_{j\geq i+1} \beta_j X_{-j}\right) \qquad a.s.$$
We obtain a contradiction using the stationarity and the point 3) considered with β_{j+i} instead of β_j, $j \geq 1$. Hence $\alpha_i = \alpha_0 = 0$, for all $i \geq 1$ and the result follows. □

From the previous Lemma we deduce the identification condition:

Lemma 6 *If* **(A3)** *holds then*
$$(m_0(\theta), V_0(\theta)) = (m_0(\theta_0), V_0(\theta_0)) \quad a.s \Rightarrow \theta = \theta_0$$

Proof. The equality $m_0(\theta) = m_0(\theta_0)$ writes as:
$$\sum_{j=1}^{q} (b_j - b_{0,j}) Y_{t-j} = 0 \quad \text{a.s.}$$
If there exists $j \in \{1, \ldots, q\}$ such that $b_j \neq b_{0,j}$, then there exists $j \in \{1, \ldots, q\}$ such that $X_{t-j} = \xi_{t-j}\sigma_{t-j}(\theta_0) \in \mathcal{F}_{t-j-1}$. Using the same argument as in the proof of point 1. in Lemma 5, we obtain $\sigma_{t-j}(\theta_0) = 0$ a.s. Then $X = 0$ a.s and this is a contradiction with $a_{0,0} \neq 0$. We deduce that $b_j = b_{0,j}$, $\forall j \in \{1, \ldots, q\}$.
Assume that equality $V_0(\theta) = V_0(\theta_0)$ a.s holds, then as
$$m_0(\theta) = m_0(\theta_0) \quad \text{a.s} \Rightarrow m_{-j}(\theta) = m_{-j}(\theta_0), \quad j = 1, \ldots, p \quad \text{a.s},$$
we obtain using equation (9):
$$\left(a_0 + a_{0,0} + \sum_{j=1}^{p} (a_j + a_{0,j}) X_{-j}\right)\left(a_0 - a_{0,0} + \sum_{j=1}^{p} (a_j - a_{0,j}) X_{-j}\right) = 0 \qquad \text{a.s}$$

As $a_0 + a_{0,0} > 0$ we obtain $a_0 = a_{0,0}$ by using point 4) in Lemma (5), and $a_j = a_{0,j}$ for all $j = 1, \ldots, p$. Thus $\theta = \theta_0$. $\Box$

Now we prove theorem 1. The proof follows the proof of theorem 2.1 in Jeantheau [11] who proved the consistency of the QMLE for general multivariate ARCH models (see theorem 2.1 of that paper). As in [11] we use the following Theorem which is a staightforward generalisation of Theorem 1.12 in Pfanzagl (1969) for i.i.d data.

Theorem 3 *Let $(Y_t)_{t\in\mathbb{Z}}$ be a strictly stationary and ergodic process, θ a parameter in Θ a compact of $\mathbb{R}^d$, and for $n \in \mathbb{N}^*$, Q_n be a contrast such that*

$$Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} f\left(Y_t, \ldots, Y_{t-p}; \theta\right)$$

where f is a measurable function with real values and continuous in θ. Suppose that
1) $\mathbb{E}\inf_{\theta\in\Theta} f\left(Y_0, \ldots, Y_{-p}; \theta\right) > -\infty$.
2) $\theta \mapsto \mathbb{E}f\left(Y_0, \ldots, Y_{-p}; \theta\right)$ has a unique finite minimum at θ_0,
The minimum contrast estimator $\hat{\theta}_n$ associated to Q_n is thus strongly consistent: $\lim_{n\to\infty}\hat{\theta}_n = \theta_0$ a.s.

We apply Theorem 3 setting $f\left(Y_0, \ldots, Y_{-p}; \theta\right) = q_{0,h}(\theta)$. Obviously f is continuous in θ.

- Since $\inf_{\theta\in\Theta} f\left(Y_0, \ldots, Y_{-p}; \theta\right) \geq \ln h$, assumption 1) of Theorem 3 holds for the AR-LARCH process Y.

- We next prove that assumption 2) holds. We first prove that $Q_h(\theta_0) = \mathbb{E}q_{0,h}(\theta_0) \in \mathbb{R}$ (from the last point we know that $Q_h(\theta_0)$ is well defined and $\in \mathbb{R}\cup\{\infty\}$. From **(A1)**, Francq and Zakoïan [10] prove that $\mathbb{E}\left|X_0\right|^s < \infty$ for some $s \in (0, 1]$ (see the proof of Theorem 4.2 in [10]). hence:

  $$Q_h(\theta_0) = 1 + \frac{1}{s}\mathbb{E}\ln\left(V_0(\theta_0) + h\right)^s \leq 1 + \frac{h^s}{s} + \mathbb{E}V_0(\theta_0)^s < \infty.$$

  Now we prove that for $\theta \in \Theta$, $Q_h(\theta) \geq Q_h(\theta_0)$ and the equality holds only when $\theta = \theta_0$.
  For $\theta \in \Theta$, we have:

  $$\mathbb{E}\left(\frac{(Y_0 - m_0(\theta))^2 + h}{V_0(\theta) + h}\right) = \mathbb{E}\left((A\xi_0 + B)^2 + C\right),$$

  where $A = (V_0(\theta)+h)^{-1/2}\sigma_0(\theta_0)$, $B = (V_0(\theta)+h)^{-1/2}(m_0(\theta_0) - m_0(\theta))$ and $C = (V_0(\theta) + h)^{-1}h$. If μ is the law of the vector (A, B, C), then we obtain using independence properties:

  $$\mathbb{E}\left((A\xi_0 + B)^2 + C\right) = \int \mathbb{E}\left((a\xi_0 + b)^2 + c\right)\mu(da, db, dc) = \mathbb{E}\left(A^2 + B^2 + C\right),$$

and we have proved that

$$Q_h(\theta) = \mathbb{E}\left(\frac{(m_0(\theta) - m_0(\theta_0))^2 + h}{V_0(\theta) + h} + \ln(V_0(\theta) + h)\right).$$

We obtain:

$$Q_h(\theta) - Q_h(\theta_0) = \mathbb{E}\frac{V_0(\theta_0) + h}{V_0(\theta) + h} - \ln\frac{V_0(\theta_0) + h}{V_0(\theta) + h} - 1$$

Since $\big(x - \ln x \geq 1, \ \forall x > 0\big)$ and $\big(x - \ln x = 1 \Leftrightarrow x = 1\big)$ we derive $Q_h(\theta_0) \leq Q_h(\theta)$ and:

$$Q_h(\theta) = Q_h(\theta_0) \Rightarrow m_0(\theta) = m_0(\theta_0), \quad V_0(\theta) = V_0(\theta_0) \quad a.s$$

From Lemma 6, we get $\theta = \theta_0$ which proves that assumption 2) of Theorem 3 holds.

Then the consistency of the smoothed QMLE follows from Theorem 3. □

7.4 Proof of Theorem 2

We use very classical arguments, the approach of Straumann [19] allows to derive an uniform law of the large numbers namely we will use:

Theorem 4 (Straumann (2006), Theorem 2.2.1, [19]) *Let $(v_t)_{t\in\mathbb{Z}}$ be a stationary ergodic sequence with values in $\mathcal{C}\left(K, \mathbb{R}^k\right)$, the space of real continuous functions on a compact $K \subset \mathbb{R}^d$. Assume $\mathbb{E}\sup_{\theta\in K}\|v_0(\theta)\| < \infty$ then:*

$$\lim_{n\to\infty}\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{t=1}^{n} v_t(\theta) - \mathbb{E}v_0(\theta)\right| = 0, \qquad a.s$$

Before we prove Theorem 2 we need the two following lemmas:

Lemma 7 *Suppose that* **A3** *holds and let $c \in \mathbb{R}^{p+q+1}$ such that $c'\nabla m_0(\theta_0) = c'\nabla V_0(\theta_0) = 0$ a.s. Then $c = 0$.*

Proof. We compute $\partial m_0/\partial b_j(\theta_0) = Y_{-j}$ for $j = 1, \ldots, p$, $\partial V_0/\partial a_0(\theta_0) = 2\sigma_0(\theta_0)$ and $\partial V_0/\partial a_j(\theta_0) = 2X_{-j}\sigma_0(\theta_0)$ for $j = 1, \ldots, p$. Suppose that there exists $c = (\mu_1, \ldots, \mu_q, \lambda_0, \ldots, \lambda_p) \in \mathbb{R}^{p+q+1}$ such that,

$$c'\nabla m_0(\theta_0) = c'\nabla V_0(\theta_0) = 0, \quad \text{a.s.}$$

Then we obtain:

$$\sum_{j=1}^{q} \mu_j Y_{-j} = 0, \quad \text{a.s.}$$

As for the proof of Lemma 6 we obtain $\mu_1 = \ldots = \mu_q = 0$.
Hence equality $c'\nabla V_0(\theta_0) = 0$ rewrites:

$$\sigma_0(\theta_0)\Big(\lambda_0 + \sum_{j=1}^{d} \lambda_j X_{-j}\Big) = 0 \qquad a.s.$$

As $a_{0,0} \neq 0$ an application of point 4) of Lemma 5 implies $\lambda_j = 0$, $j = 0, \ldots, p$. We have shown that $c = 0$. □
For the proof of theorem 2, the following moment conditions will be used:

Lemma 8 *If* $\mathbb{E}X_0^4 < \infty$ *then* $\mathbb{E}\left\|\nabla q_{0,h}(\theta_0)\right\|^2 < \infty$, $\mathbb{E}\sup_{\theta\in\Theta}\left\|\nabla^2 q_{0,h}(\theta)\right\| < \infty$.

Proof. We first notice that

$$\nabla q_{0,h}(\theta) = \frac{-2(Y_0 - m_0(\theta))\nabla m_0(\theta)}{V_0(\theta) + h} + \frac{\nabla V_0(\theta)\left(V_0(\theta) - (Y_0 - m_0(\theta))^2\right)}{(V_0(\theta) + h)^2} \quad (23)$$

$$\begin{aligned}
\nabla^2 q_{0,h}(\theta) &= \frac{2}{V_0(\theta)+h}\nabla m_0(\theta)\nabla m_0(\theta)' \\
&- \frac{2(Y_0 - m_0(\theta))}{V_0(\theta)+h}\nabla^2 m_0(\theta) \\
&+ \frac{2(Y_0 - m_0(\theta)}{(V_0(\theta)+h)^2}\left(\nabla m_0(\theta)\nabla V_0(\theta)' + \nabla V_0(\theta)\nabla m_0(\theta)'\right) \\
&+ \frac{V_0(\theta) - (Y_0 - m_0(\theta))^2}{(V_0(\theta)+h)^2}\nabla^2 V_0(\theta) \\
&+ \frac{h - V_0(\theta) + 2(Y_0 - m_0(\theta))^2}{(V_0(\theta)+h)^3}\nabla V_0(\theta)\nabla V_0(\theta)'
\end{aligned}$$

As Θ is bounded since it is compact, the expressions of σ_0 and m_0 for model 9 entails the existence of a real $K > 0$ such that:

$$\sup_{\theta\in\Theta}\left(|m_0(\theta)| + \|\nabla m_0(\theta)\| + |\sigma_0(\theta)| + \|\nabla\sigma_0(\theta)\| + \left\|\nabla^2\sigma_0(\theta)\right\|\right) \leq U, \quad (24)$$

with $U = K\left(1 + \sum_{j=1}^{p+q}|Y_{-j}|\right)$. Moreover for model (9), $\nabla^2 m_0 = 0$.

- For $\nabla q_{0,h}$, we have:

$$\nabla q_{0,h}(\theta_0) = \frac{-2X_0\nabla m_0(\theta_0)}{V_0(\theta_0)+h} + \frac{\nabla V_0(\theta_0)(1-\xi_0^2)^2 V_0(\theta_0)}{(V_0(\theta_0)+h)^2}.$$

  Then,

$$\|\nabla q_{0,h}(\theta_0)\|^2 \leq \frac{4X_0^2U^2}{(V_0(\theta_0)+h)^2} + \frac{V_0(\theta_0)^3U^2(1-\xi_0^2)^2}{(V_0(\theta_0)+h)^4}$$

This leads to:

$$\mathbb{E}\left\|\nabla q_{0,h}(\theta_0)\right\|^2 \leq \frac{3+\mu_4}{h}\mathbb{E}U^2.$$

As $\mathbb{E}X_0^4 < \infty \Rightarrow \mathbb{E}Y_0^4 < \infty \Rightarrow \mathbb{E}U^2 < \infty$, we obtain $\mathbb{E}\left\|\nabla q_{0,h}(\theta_0)\right\|^2 < \infty$.

- For the second assertion, using the definition of U and the inequality $\frac{1}{V_0(\theta)+h} \leq \frac{1}{h}$, we see that the fourth first term in the expression of $\nabla^2 q_{0,h}(\theta)$ can be bounded uniformly with respect to θ by polynomials of degree four in the variables $|Y_0|, |Y_{-1}|, \ldots, \left|Y_{-(p+q)}\right|$. For the last term it is also the case since:

$$\begin{aligned}
&\frac{\left|h - V_0(\theta) + 2(Y_0 - m_0(\theta))^2\right|}{(V_0(\theta)+h)^3}\left\|\nabla V_0(\theta)\right\|^2 \\
&\qquad \leq \frac{h + U^2 + 2(|Y_0| + U)^2}{(V_0(\theta)+h)^3} 4V_0(\theta)U^2 \\
&\qquad\qquad \leq \frac{h + U^2 + 2(|Y_0| + U)^2}{h^2} 4U^2.
\end{aligned}$$

The result follows thus from $\mathbb{E}Y_0^4 < \infty$. $\square$

We now turn to the proof of Theorem 2. Since $\theta \in \Theta^\circ$, a Taylor expansion yields:

$$0 = \nabla Q_{n,h}(\hat{\theta}_{n,h}) = \nabla Q_{n,h}(\theta_0) + \widetilde{M}_n \cdot (\hat{\theta}_{n,h} - \theta_0),$$

with $\widetilde{M}_n(i,j) = \partial^2 Q_{n,h}(\gamma_i)/\partial\theta_i\partial\theta_j$, and $\left\|\hat{\theta}_{n,h} - \gamma_i\right\| \leq \left\|\hat{\theta}_{n,h} - \theta_0\right\|$, for $i = 1, \ldots, p+q+1$.

Hence,

$$-\sqrt{n}\nabla Q_{n,h}(\theta_0) = \sqrt{n}\,\widetilde{M}_n \cdot (\hat{\theta}_{n,h} - \theta_0) \tag{25}$$

For each $(\theta, t) \in \Theta \times \mathbb{Z}$, $\nabla^2 q_{t,h}(\theta)$ is a measurable function of $Y_t, \ldots, Y_{t-(p+q)}$, thus we infer that $\left(\nabla^2 q_{t,h}\right)_t$ is a stationary ergodic and $\mathcal{C}(\Theta, \mathbb{R}^{p+q+1} \times \mathbb{R}^{p+q+1})$-valued sequence. According to Lemma 8, $\sup_{\theta\in\Theta}\left\|\nabla^2 q_{0,h}(\theta)\right\|$ is an integrable random variable, hence from Theorem 4:

$$\sup_{\theta\in\Theta}\left\|\nabla^2 Q_{n,h}(\theta) - \mathbb{E}\nabla^2 q_{0,h}(\theta)\right\| \to_{n\to\infty} 0.$$

From $\hat{\theta}_{n,h} \to_{n\to\infty} \theta_0$ *(a.s.)* we thus conclude

$$\widetilde{M}_n \to_{n\to\infty} N_h = \mathbb{E}\nabla^2 q_{0,h}(\theta_0), \qquad a.s. \tag{26}$$

Moreover N_h is non-singular; indeed using expression (23), we have

$$N_h = 2\mathbb{E}\frac{\nabla m_0(\theta_0)\nabla m_0(\theta_0)'}{V_0(\theta_0)+h} + \mathbb{E}\frac{\nabla V_0(\theta_0)\nabla V_0(\theta_0)'}{(V_0(\theta_0)+h)^2},$$

and from Lemma 7 this matrix is positive-definite. Now,

$$\sqrt{n}\,\nabla Q_{n,h}(\theta_0) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\nabla q_{n,h}(\theta_0), \quad \text{with } \mathbb{E}_{\mathcal{F}_{t-1}}\nabla q_0(\theta_0) = 0.$$

Since from Lemma 8, $\mathbb{E}\left\|\nabla q_{0,h}(\theta_0)\right\|^2 < \infty$, $\left(\nabla q_{t,h}(\theta_0)\right)_t$ is an ergodic stationary $\mathcal{F}_t$-martingale difference sequence with finite variance and from Theorem 23.1, page 206 in [4],

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\nabla q_{t,h}(\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, M_h), \quad \text{with } M_h = \operatorname{Var}\nabla q_{0,h}(\theta_0).$$

Thus we infer

$$\sqrt{n}(\hat{\theta}_{n,h} - \theta_0) \rightarrow_{n\to\infty} \mathcal{N}(0, N_h^{-1}M_hN_h^{-1})$$

Then the result follows from the expression of M_h which is easy to derive from the expression of $\nabla q_{0,h}$ (23) . $\square$

7.5 Proof of Lemma 3

We use the expression of v_h given in Theorem 2. We will only prove the Lemma if $q \neq 0$ and $(\mu_3, \mu_4) = (0, 3)$. The case $q = 0$ is omitted since its proof follows from straightforward modifications.
If $(\mu_3, \mu_4) = (0, 3)$ we first remark that $M_h \preceq 2N_h$ and we have the following bound:

$$v_h \preceq 2N_h^{-1} \tag{27}$$

We now prove the Lemma.

1. If $y, z \in \mathbb{R}^d$ and $h, k > 0$ we have:

$$\begin{aligned}\sqrt{2}\,y'N_k^{(1)}z &= 2\sqrt{2}\,\mathbb{E}\frac{(m/y)\times(m/z)}{V+k}\\ &\leq 2\sqrt{2}\,\mathbb{E}^{1/2}\left(\frac{(m/y)^2(V+h)}{(V+k)^2}\right)\times\mathbb{E}^{1/2}\left(\frac{(m/z)^2}{V+h}\right)\end{aligned}$$

   With analogous arguments we also have:

$$\sqrt{2}\,y'N_k^{(2)}z \leq \sqrt{2}\,\mathbb{E}^{1/2}\left(\frac{(V/y)^2(V+h)^2}{(V+k)^4}\right)\times\mathbb{E}^{1/2}\left(\frac{(V/z)^2}{(V+h)^2}\right)$$

   Now using from the inequality $(ac+bd)^2 \leq (a^2+b^2)(c^2+d^2)$:

$$2\left(y'N_kz\right)^2 \leq y'M_{k,h}y \times z'N_hz,$$

where $M_{k,h} = 4\mathbb{E}\frac{\nabla m \nabla m'(V+h)}{(V+k)^2} + 2\mathbb{E}\frac{\nabla V \nabla V'(V+h)^2}{(V+k)^4}$.
Now if $z = N_h^{-1}x$ and $y = N_k^{-1}x$, we get:

$$2x'N_h x \leq x'N_k^{-1}M_{k,h}N_k^{-1}x.$$

Since $\lim_{h\to 0} M_{k,h} = M_k$, we obtain using (27):

$$\limsup_{h\to 0^+} x'v_h x \leq 2\limsup_{h\to 0^+} x'N_h^{-1}x \leq x'v_k x. \tag{28}$$

We deduce that

$$\limsup_{h\to 0^+} x'v_h x \leq \liminf_{k\to 0^+} x'v_k.$$

The last inequality is obviously an equality and we conclude that for $x \in \mathbb{R}^{p+q+1}$, $\lim_{h\to 0} x'v_h x$ exists and belongs to $\mathbb{R}^+$. By polarization $\lim_{h\to 0} x'v_h y$ exists for all $x, y \in \mathbb{R}^{p+q+1}$. Then one can define $v = \lim_{h\to 0^+} v_h$. From (28), we deduce $v \preceq v_k$ if $k > 0$.

2. Suppose first that $\mathcal{C}$ holds. Then from the dominated convergence theorem we prove that

$$\lim_{h\to 0} M_h = 2\lim_{h\to 0} N_h = 2N.$$

From Lemma 7 this limit is nondegenerated. The expression of v in this case follows now from the continuity of the application $A \mapsto A^{-1}$. Now, suppose that $\mathcal{C}$ does not hold. We need to prove that v is degenerated. From the previous points, $v \preceq v_h \preceq 2N_h^{-1}$ for $h > 0$. Let λ_h be the smallest eigenvalue of the matrix N_h^{-1}, then for each $h > 0$:

$$0 \leq \inf_{\|x\|=1} x'vx \leq \inf_{\|x\|=1} 2x'N_h^{-1}x \leq 2\lambda_h.$$

To prove that v is degenerated this is enough to show that $\lim_{h\to 0^+} \lambda_h = 0$. As $\lambda_h = 1/\rho(N_h)$ where $\rho(N_h)$ denotes the spectral radius of the matrix N_h we need to show that $\lim_{h\to 0^+} \rho(N_h) = \infty$ or equivalently that $\lim_{h\to 0^+} \sum_{i=1}^{p+q+1} N_h(i,i) = \infty$. But

$$\sum_{i=1}^{p+q+1} N_h(i,i) = 2\mathbb{E}\|m\|^2(V+h)^{-1} + \mathbb{E}\|V\|^2(V+h)^{-2}$$

and with monotone convergence:

$$\lim_{h\to 0^+} 2\mathbb{E}\frac{\|m\|^2}{V+h} + \mathbb{E}\frac{\|V\|^2}{(V+h)^2} = 2\mathbb{E}\frac{\|m\|^2}{V}\mathbf{1}_{V\neq 0} + \mathbb{E}\frac{\|Y\|^2}{V^2}\mathbf{1}_{V\neq 0} = \infty.$$

Hence we conclude that v is degenerated. □

7.6 Proof of Lemma 4

If $a_{0,1}^4\mathbb{E}\xi_0^4 < 1$ then $a_{0,1}^2 < 1$ and from Theorem 2.1 in [12] there exists a unique stationary solution of equation (18). The fourth moment of this solution exists from (7).

If now there exists a stationary solution of equation (18) such that $\mathbb{E}X_0^4 < \infty$. As

$$\mathbb{E}X_0^4 = \mathbb{E}\xi_0^4\mathbb{E}\left(a_{0,0}^4 + a_{0,1}^4X_0^4 + 6a_{0,0}^2a_{0,1}^2X_0^2 + 4a_{0,0}a_{0,1}^3X_0^3 + 4a_{0,0}^3a_{0,1}X_0\right),$$

since $\mathbb{E}\xi_0^3 = 0$ implies $\mathbb{E}X_0^3 = 0$ we get:

$$\mathbb{E}X_0^4 = \mathbb{E}\xi_0^4\mathbb{E}\left(a_{0,0}^4 + a_{0,1}^4X_0^4 + 6a_{0,0}^2a_{0,1}^2X_0^2\right).$$

Hence $\left(1 - a_{0,0}^4\mathbb{E}\xi_0^4\right)\mathbb{E}X_0^4 = \mathbb{E}\xi_0^4\left(a_{0,0}^4 + 6a_{0,0}^2a_{0,1}^2\mathbb{E}X_0^2\right)$ and $a_{0,1}^4\mathbb{E}\xi_0^4 < 1$. $\square$

References

[1] Bardet, J.-M., Wintenberger, O. (2007) *Asymptotic normality of the quasi maximum likelihood estimator for multidimensional causal processes.* Preprint `arXiv:0712.0679`.

[2] Black, F. (1976) *Studies in stock price volatility changes.* Proceedings of the Business and Economic Statistics Section, 177-181.

[3] Berkes, I., Horváth, L., Kokoszka, P. S. (2003) *GARCH processes: structure and estimation.* Bernoulli **9**, 201-227.

[4] Billingsley, P. (1968) *Convergence of Probability Measures.* John Wiley, New York.

[5] Doukhan, P., Oppenheim, G., Taqqu, M., editors (2003) *Theory and Applications of Long-range Dependence.* Birkhaüser, Boston.

[6] Doukhan, P., Teyssière, G., Winant, P. (2006) *Vector valued ARCH infinity processes.* Dependence in Probability and Statistics. Patrice Bertail, Paul Doukhan and Philippe Soulier Editors, Springer, New York.

[7] Engle, R. F. (1982) *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation.* Econometrica **50**, 987-1008.

[8] Engle, R. F. (1990) *Stock volatility and the crash of 87. Discussion.* The Review of Financial Studies **3**, 103-106.

[9] Francq, C., Zakoïan, J-M. (2004) *Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes.* Bernoulli, 10, 605-637.

[10] Francq, C., Makarova, S., Zakoïan, J-M. (2008) *A class of stochastic unit-root bilinear processes. Mixing properties and unit-root test.* Forthcoming in the Journal of Econometrics.

[11] Jeantheau, T. (1998) *Strong consistency of estimators for multivariate ARCH models.* Econometric Theory **14**-1, 70-86.

[12] Giraitis, L., Robinson, P., Surgailis, D. (2000) *A model for long memory conditional heteroscedasticity.* Annals of Applied Probability **10**-3, 1002-1024.

[13] Giraitis, L., Leipus, R., Robinson, P., Surgailis, D. (2004) *LARCH, Leverage, and Long Memory.* Journal of Financial Econometrics **2**-2, 177-210.

[14] Giraitis, L., Surgailis, D. (2002) *ARCH-type bilinear models with double long memory.* Stochastic Processes and their Applications **100**, 275-300.

[15] Lee, S.-W., Hansen, B. E. (1994) *Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator.* Econometric theory **10**, 29-52.

[16] Lumsdaine, R. L. (1996) *Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models.* Econometrica **64**, 575-596.

[17] Mikosch, T., Straumann, D. (2006) *Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach.* Annals of Statistics **34**-5, 2449-2495.

[18] Sentana, E. (1995) *Quadratic ARCH models.* Review of Economic Studies **62**, 639-661.

[19] Straumann, D. (2004) *Estimation in Conditionally Heteroscedastic Time Series Models.* Lecture Notes in Statistics. Springer Verlag.

[20] Weiss, A. A. (1986) *Asymptotic theory for ARCH models: estimation and testing.* Econometric theory **2**, 107-131.

LS-CREST, Laboratoire de Statistique, CREST. Timbre J340, 3, avenue Pierre Larousse, 92240 Malakoff, and
SAMOS, Statistique Appliquée et MOdélisation Stochastique, Université Paris 1, Centre Pierre Mendès France, 90 rue de Tolbiac, F-75634 Paris Cedex 13, France. `Email: truquet@ensae.fr.`